

SNS COLLEGE OF TECHNOLOGY

**An Autonomous Institution
Coimbatore-35**



DEPARTMENT OF ARTIFICIAL INTELLIGENCE & DATA SCIENCE

23ADT202 – FUNDAMENTALS OF DATA SCIENCE AND ANALYTICS

II YEAR IV SEM

UNIT III – INFERENCE STATISTICS

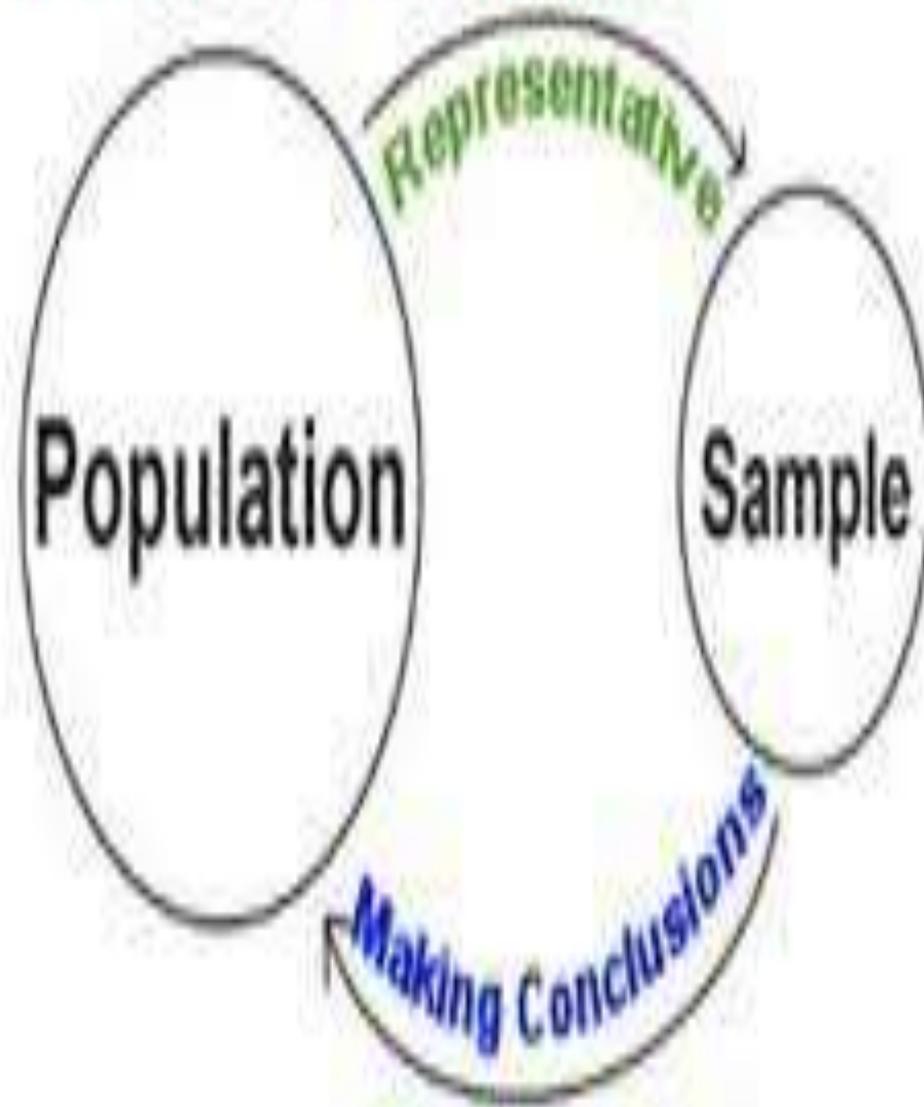
POPULATION AND SAMPLES

What is Inferential Statistics?

EMPATHY:

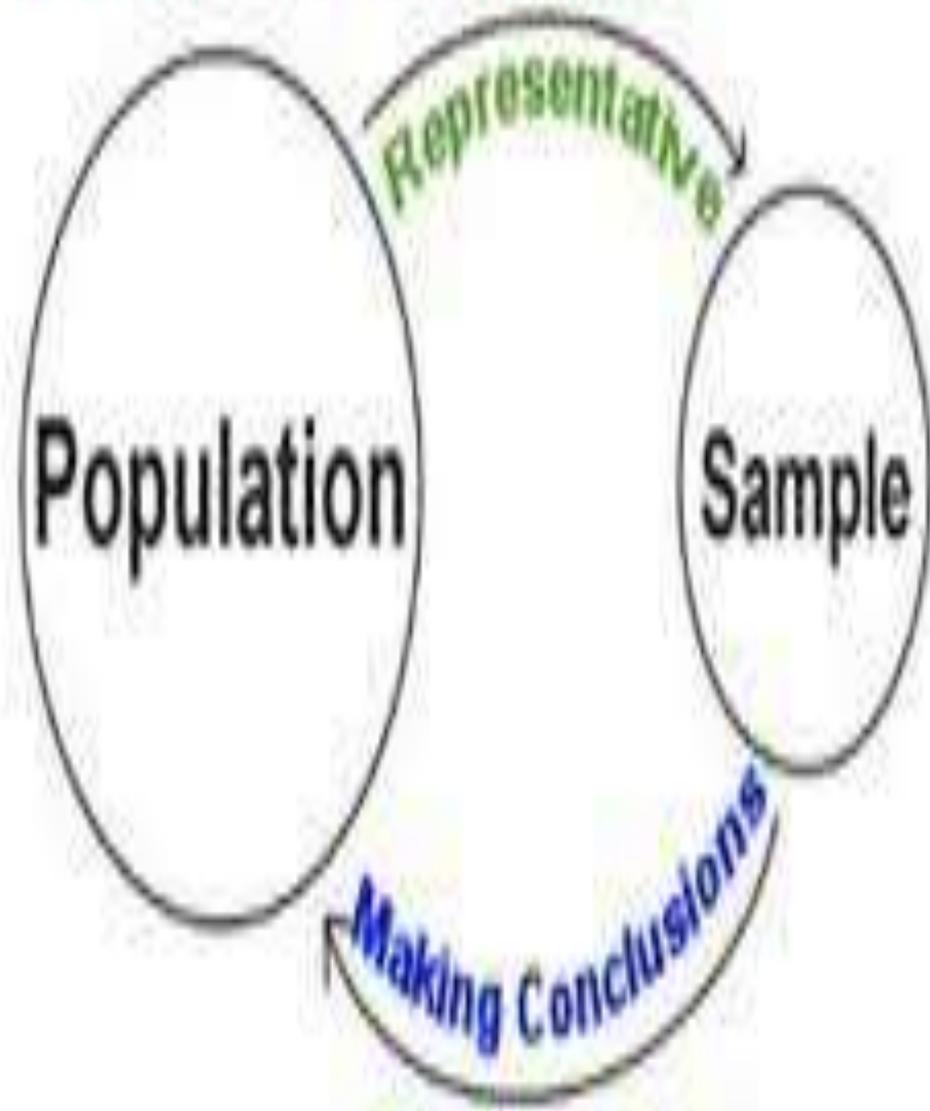
- ❖ Inferential statistics is an **important tool** that allows us to make **predictions and conclusions** about a population based on sample data.
- ❖ Unlike descriptive statistics, which only **summarize data**, inferential statistics let us **test hypotheses**, **make estimates**, and **measure the uncertainty** about our predictions.
- ❖ These tools are essential for **evaluating models**, testing assumptions, and supporting data-driven decision-making.

Inferential Statistics

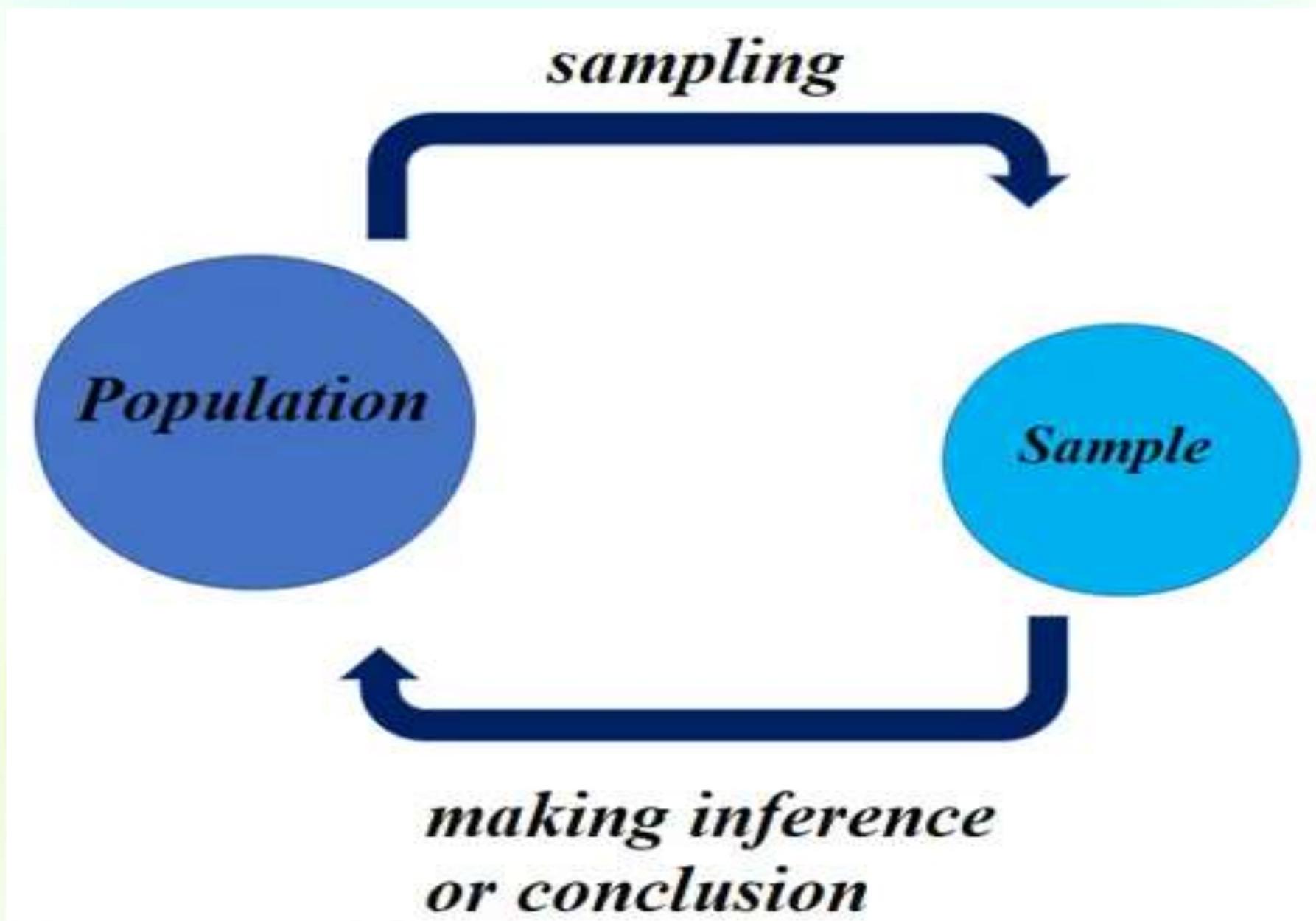


What is Inferential Statistics?

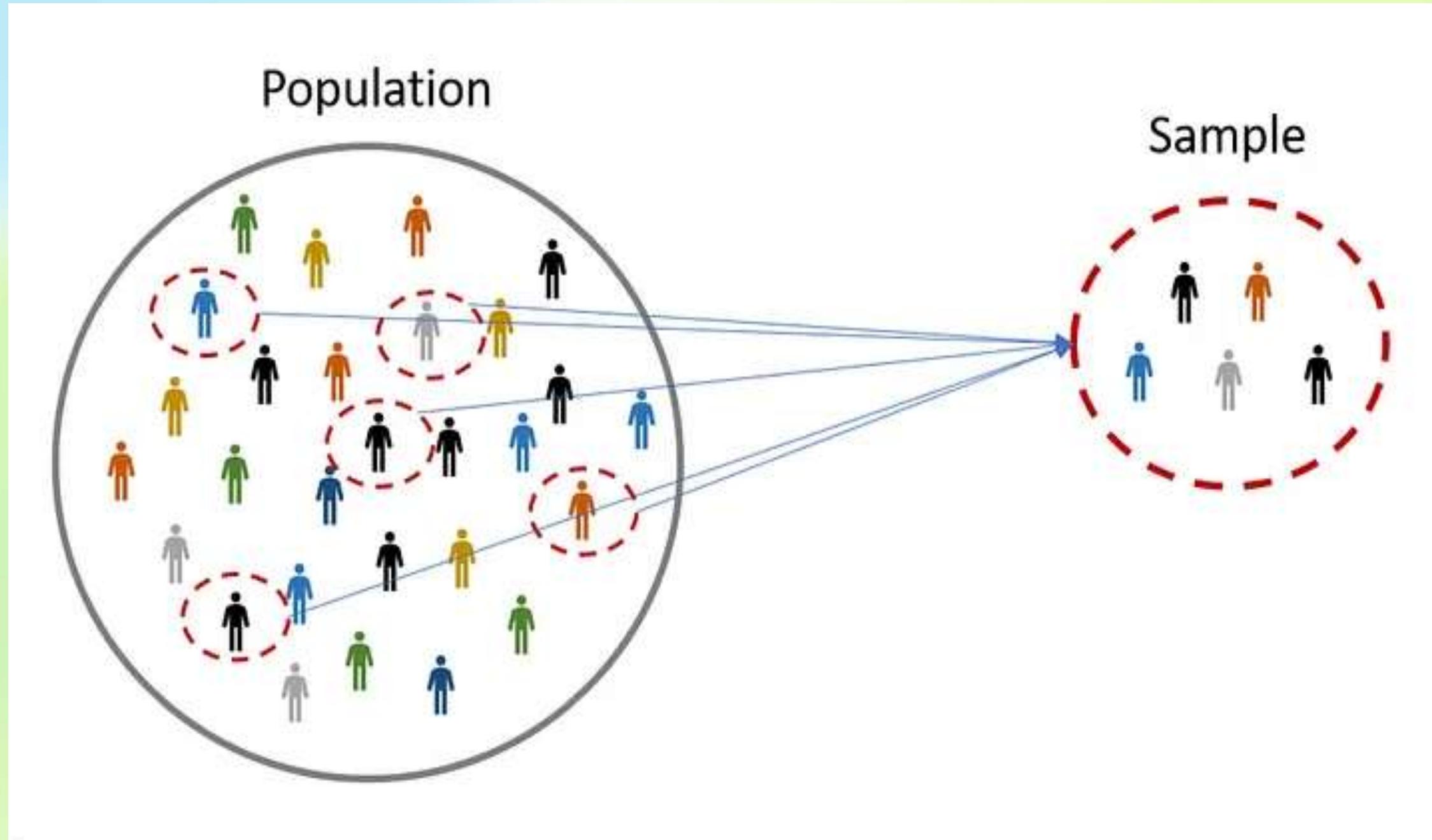
Inferential Statistics

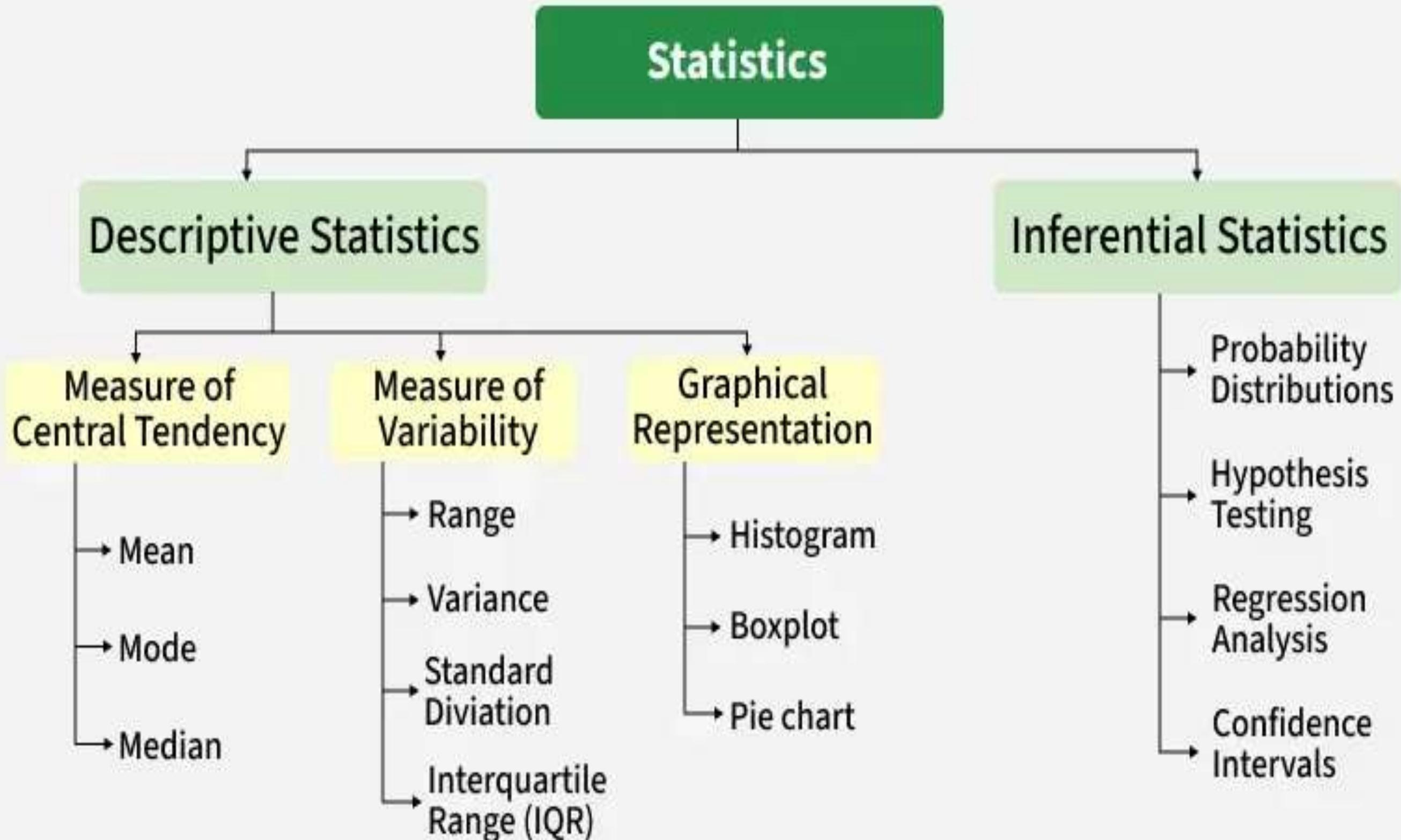


- *For example, instead of surveying every voter in a country, we can survey a few thousand and still make reliable conclusions about the entire population's opinion.*
- *Inferential statistics provides the tools to do this **systematically and mathematically**.*



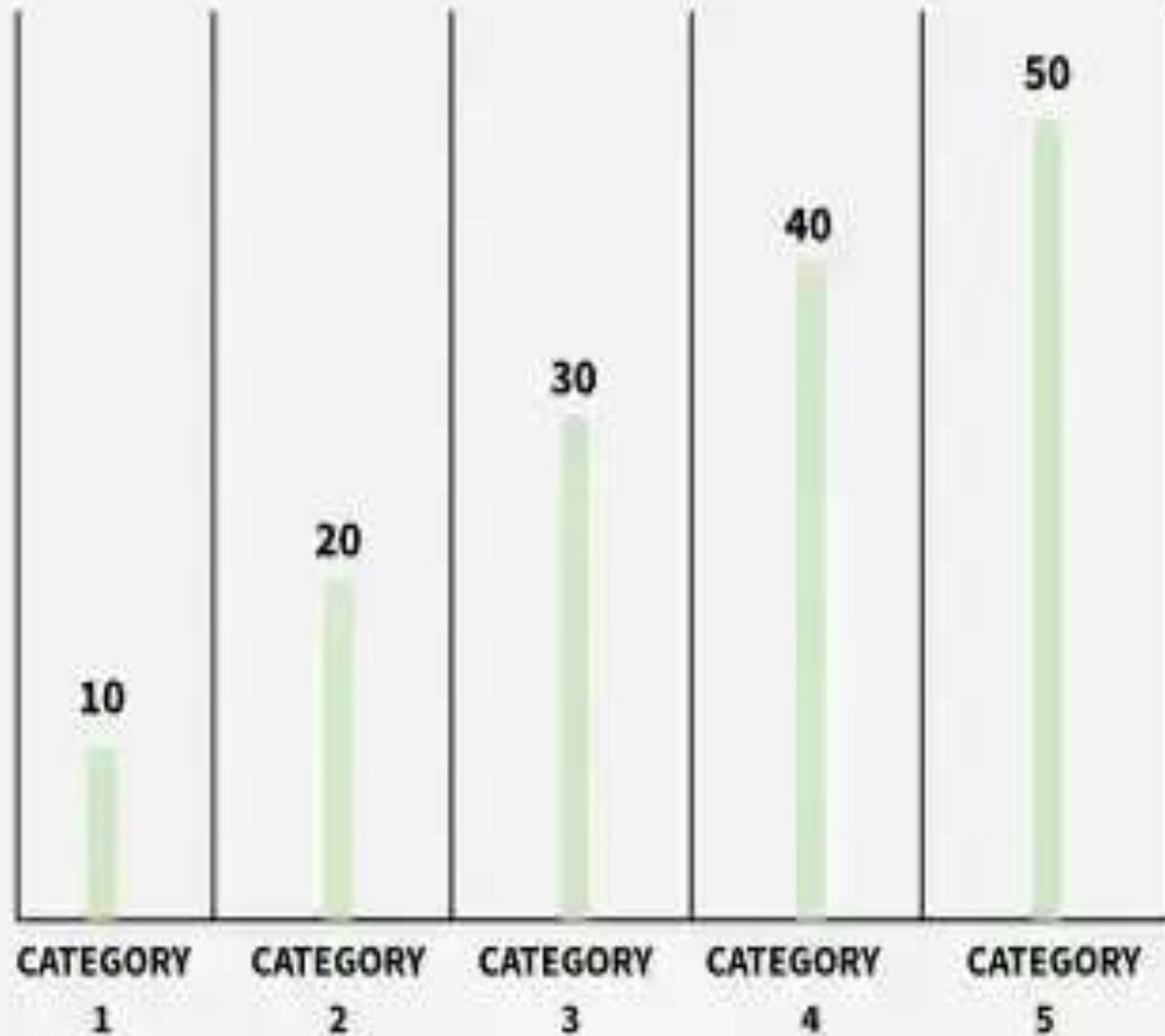
Population-Sample





Descriptive Statistics

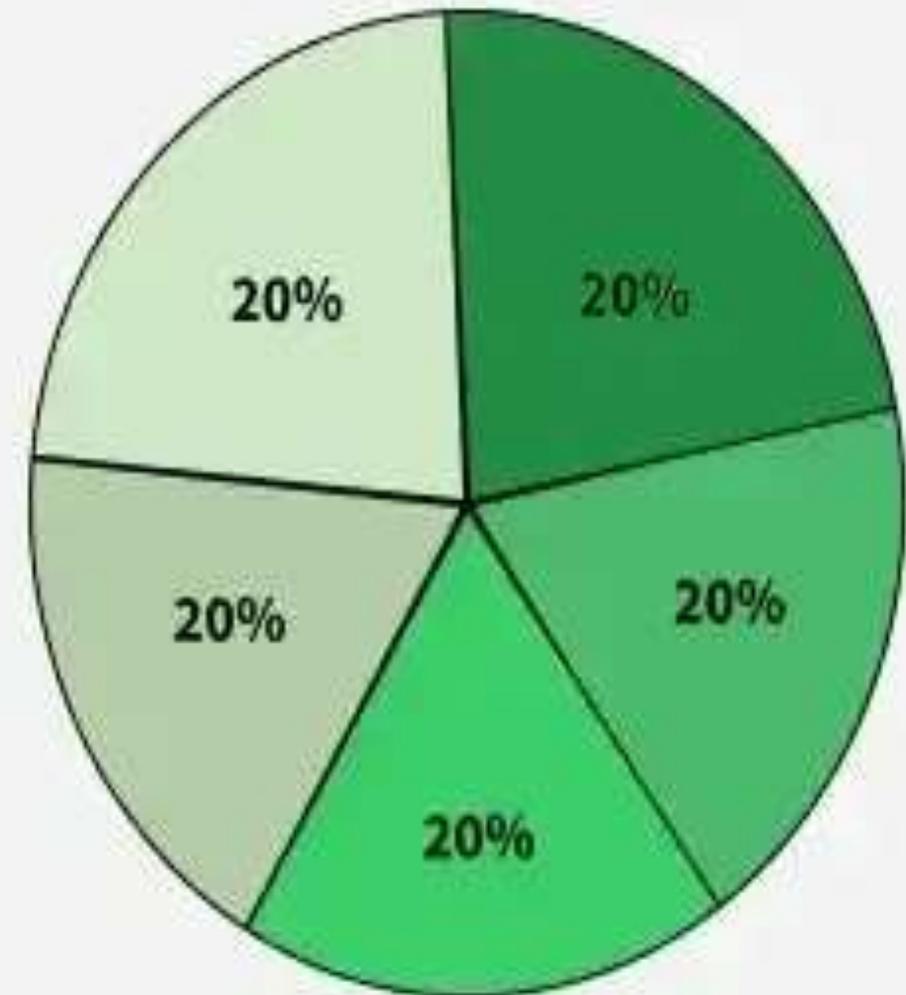
Describes the characteristics of data
 Example: Population, Frequency of the variables



Vs

Inferential Statistics

Studies the sample of the same data
 Example: Grade, Percentile



DEFINE:**Why Do We Need Inferential Statistics?**

In real-world scenarios, **analyzing an entire population is often impossible.**

Instead, we collect data from a sample and use inferential statistics to:

- Conclude the whole population.
- Test claims or hypotheses.
- Calculate confidence intervals and p-values to measure uncertainty.
- Make predictions with statistical models.

Population:

Population : The Population is the **Entire group** that we are taking for analysis or prediction.

For instance, a **population** might be the set of:

- All students at a university
- All the cell phones ever manufactured by a company
- All the forests on Earth

Sample :

Sample : Sample is the **Subset of the Population** (i.e. Taking random samples from the population). **The size of the sample is always less than the total size of the population.**

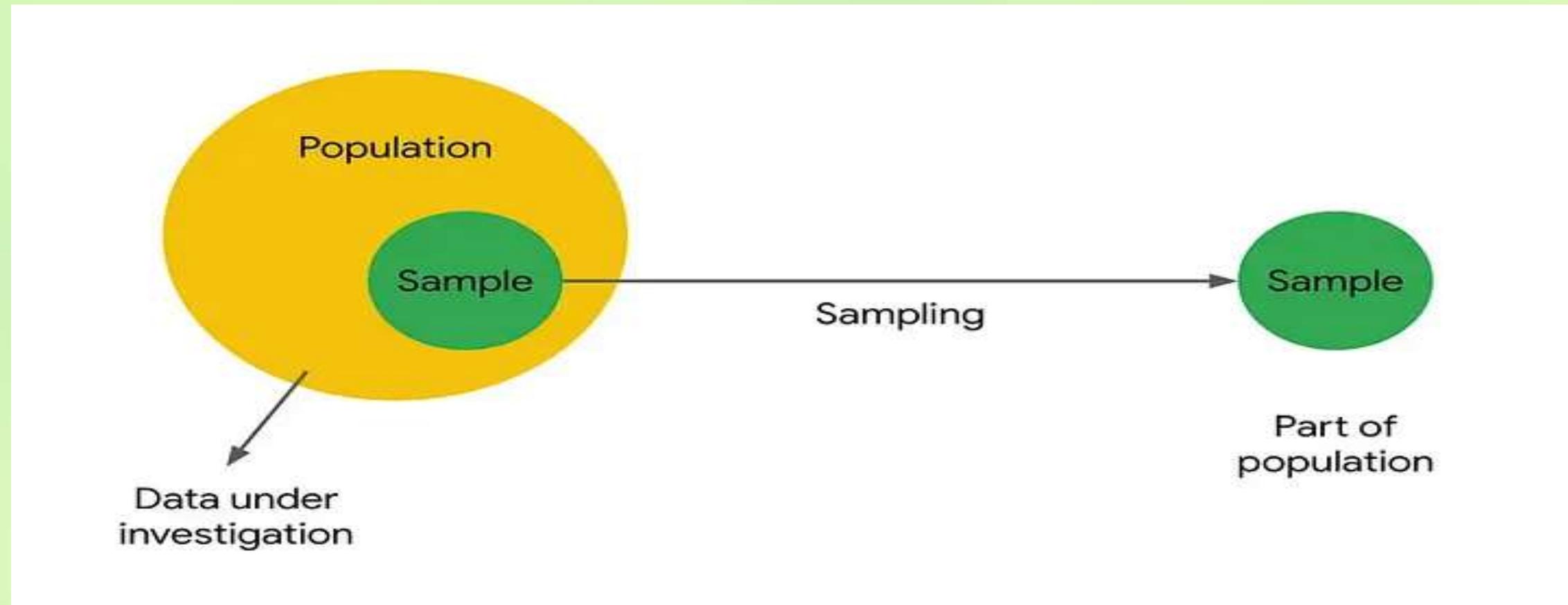
A **sample** is a subset of a population.

Samples drawn from the above populations might be:

- The math majors at the university
- The cell phones manufactured by the company in the last week
- The forests in Canada

Sample :

- Data professionals use samples to **make inferences** about populations.
- In other words, they use the data they collect from a small part of the population to **draw conclusions** about the population as a whole.



PROTOTYPING:

Step-by-Step Process:

1. Define the Population and Sample:

- **Population:** The entire pot of soup.
- **Sample:** The spoonful of soup you tasted.

2. Collect Data:

- You taste a spoonful of soup and assess its flavor.

3. Assumption and Inference:

- Based on the taste of the spoonful (sample), you make an inference about the taste of the entire pot of soup (population).
- Assuming the entire pot has a similar taste based on the sample.

TESTING:

Step-by-Step Process:

4. Inferential Statistics in Action:

- **Sample Mean (\bar{x}):**

The taste quality you experience from the spoonful. Let's say you rate it as "tasty."

- **Population Mean (μ):**

The inferred taste quality of the entire pot of soup. You assume it is also "tasty" based on your sample.